

Conference Abstract

# Assessment of the quality of accession describing metadata on plant genetic resources

Markus Oppermann<sup>‡</sup>, Stephan Weise<sup>‡</sup>

<sup>‡</sup> Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

Corresponding author: Markus Oppermann ([opperman@ipk-gatersleben.de](mailto:opperman@ipk-gatersleben.de))

Received: 29 Mar 2019 | Published: 30 Jul 2019

Citation: Oppermann M, Weise S (2019) Assessment of the quality of accession describing metadata on plant genetic resources. Biodiversity Information Science and Standards 3: e34973. <https://doi.org/10.3897/biss.3.34973>

## Abstract

In the wide-ranging field of biodiversity conservation, genebanks play a major role in the preservation of cultivated plants. An important focus of genebanks is the comprehensive documentation of the maintained material. This is a prerequisite to enable users to select the most suitable material for e.g. research or breeding programs (Hoisington et al. 1999). The German Federal *ex situ* Genebank for Agricultural and Horticultural Crops, which is being hosted at IPK, is the largest genebank in Western Europe.

Within the multitude of data associated with plant material (e.g. from various -omics areas or conservation management), the so-called passport data represent the most original and oldest data in genebanks. These metadata are often subject to heterogeneity due to historically different collection and curation, especially if they were received from different institutions around the world. This leads to difficulties in handling these data and can result in misinterpretations. In addition, there are correlations between the individual attributes of the passport data which can lead to a different importance of the individual data points for the users.

Major challenges for users are to estimate completeness, correctness and reliability of these data. Thus, it is necessary to assess the quality of these data by defining a suitable set of metrics. Unfortunately, classical data quality measurement metrics, e.g. (Klier 2008), are not sufficient to fulfill the users' needs. Depending on the intention of the user, a different focus is placed on the data. Moreover, the individual attributes of the respective

areas can be related to each other. Therefore, a single index value for estimating the quality of a passport record is not sufficient. Rather, it seems to be more promising to generate more differentiated quality statements.

We are working on a metrics system that is sensitive to the users' focus. Through a practical set of rules of data quality metrics for accession-related data, the user will be able to influence the weighting of individual domains (e.g. geographical origin, biological status) according to their context (fit-for-use index).

The presentation will discuss the background and will give an overview of the progress of this research activity.

## Presenting author

Markus Oppermann

## References

- Hoisington D, Khairallah M, Reeves T, Ribaut J-M, Skovmand B, Taba S, Warburton M (1999) Plant genetic resources: What can they contribute toward increased crop productivity? *Proceedings of the National Academy of Sciences* 96 (11): 5937-5943. <https://doi.org/10.1073/pnas.96.11.5937>
- Klier M (2008) Metriken zur Bewertung der Datenqualität – Konzeption und praktischer Nutzen. *Informatik-Spektrum* 31 (3): 223-236. <https://doi.org/10.1007/s00287-007-0206-0>